

### Magnitude-Messung in Umfragen: Kontexteffekte und Methode

Wegener, Bernd

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Wegener, B. (1980). Magnitude-Messung in Umfragen: Kontexteffekte und Methode. *ZUMA Nachrichten*, 4(6), 4-40.  
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210670>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

## MAGNITUDE-MESSUNG IN UMFRAGEN: KONTEXTEFFEKTE UND METHODE\*

Die Beurteilung von kategorischen "Skalen" in Umfragen erfolgt in der Regel nach intuitiven Kriterien: Randverteilungen werden inspiziert, Ausfälle aufaddiert, der Zeitaufwand im Interview spielt eine wichtige Rolle und - post factum - zählt, ob die Meßwerte in der Auswertung zu inhaltlich brauchbaren Ergebnisse führen, was in der Regel wohl heißt: Je höher die Korrelationen, desto besser. In gründlich durchgeführten Untersuchungen erhalten auch die Interviewer noch ein Mitspracherecht, zumindest im Prestetstadium; sie sollen sich zum "Schwierigkeitsgrad" einer Frageformulierung äußern und die Angemessenheit der dazugehörigen Antwortvorgaben beurteilen. All dieses zusammen mündet dann ein in eine "Skalen"-Bewertung, für die - im positiven Fall - der Jargon die Phrase bereit hält: "Die Skala läuft".

Es wäre ganz aussichtslos, auf eine Präzisierung dieses saloppen Sprachgebrauchs zu dringen, weil die Kriterien, an denen die Qualität einer Kategoriemessung in der Umfrageforschung ausgewiesen wird, unpräzise sind. Sie sind nicht fähig, Dichotomien zu erzeugen und diese konkrete Messung als "gut" und jene als "schlecht" auszuzeichnen. D.h. es gibt in diesem Bereich keine verbindlichen und harten Zurückweisungsregeln - zumindest keine solche Regeln, die forschungspraktisch, und das heißt unabhängig von langwierigen Eichungen oder gar meßstrukturellen Analysen auch anwendbar wären.

---

\* Unter dem Titel Einstellungsmessung in Umfragen: Kategorische vs. Magnitude-Skalen haben wir in den ZUMANACHRICHTEN 3 über erste Resultate bei der Verwendung von Magnitude-Skalierungen in Umfragen berichtet. Das Interesse an dieser Methode hat in der Zwischenzeit stark zugenommen und zu einer Reihe von Untersuchungen geführt, in denen mit Magnitude-Techniken gearbeitet wird. Angesichts dieses Interesses ist es an der Zeit, auf eine technische Standardisierung der Erhebungsprozeduren hinzuwirken. Das Magnitude-Verfahren erweist sich nämlich als fehleranfällig, wenn Erhebungsformate benutzt werden, die nicht den empirisch ausgetesteten Vorlagen entsprechen. Außerdem ist ein besonderes Augenmerk auf solche interview-spezifischen Randbedingungen zu richten, die sich auf die Meßresultate negativ auswirken können. Der nachstehende Forschungsbericht verfolgt das Ziel, potentiellen Anwendern der Methode unter beiden Gesichtspunkten Hilfestellung zu geben. Sein ausschließlicher Anwendungscharakter bringt es mit sich, daß theoretische Überlegungen zur Magnitude-Methode fast ganz ausgespart bleiben. Stattdessen werden in einem Anhang einige bibliographische Hinweise gegeben. Verfasser des Berichts ist Bernd Wegener.

---

Nachträgliche "Skalenanalysen" direkter Einstellungsmessungen sind darüber hinaus immer auf Aggregationen angewiesen und damit ist ausgeschlossen, daß sich in Bezug auf eine individuelle Messung und in Bezug auf den einzelnen Befragten entscheiden läßt, wie gut oder glaubwürdig seine Angaben unter meßtheoretischen Gesichtspunkten eigentlich sind.

Mit Magnitude-Messungen von subjektiven Intensitäten ist man da von vorneherein in einer besseren Position, zumindest dann, wenn die Messungen multimodal, d.h. mit mindestens zwei Reaktionsmodalitäten vorgenommen werden. Für diese Fälle ist nämlich ein stringentes Gütekriterium angebbar, anhand dessen sich aus jeder konkreten und individuellen Messung ein Index bestimmen läßt, der Auskunft über die Qualität der Skalenwerte gibt. Dieses Kriterium ist rein formal. Es leitet sich aus der besonderen Aufgabenstellung her, die Befragte bei der Anwendung einer Magnitude-Messung erfüllen müssen. Die Methode verlangt nämlich von den Befragten die Bildung impliziter Verhältnisse zwischen Reizpaaren. Wenn diese Verhältnis-schätzungen in Bezug auf ein und dieselbe Stimulusserie unabhängig voneinander mit zwei unterschiedlichen Reaktionsmodalitäten vorgenommen werden, ergibt sich mathematisch zwischen den Meßreihen beider Reaktionsmodalitäten eine Potenzfunktion, bzw. in umgekehrter Argumentation: Nur wenn eine intermodale Potenzfunktion nachweisbar ist, haben die Befragten in ihren Urteilen auch die Instruktion befolgt und tatsächlich Verhältnisschätzungen vorgenommen. Aus diesem Grund stellt die Analyse der Funktionsform zwischen den benutzten Reaktionsmodalitäten ein formales und inhaltsunabhängiges Kriterium für die Bestimmung der erreichten Skalenqualität zur Verfügung. (Im Abschnitt 1.1 werden hierzu einige genauere Bemerkungen gemacht.)

Bei mehrmodalen Magnitude-Messungen ist somit parallel zur Konstruktion individueller Meßwerte immer auch deren Güte überprüfbar. Als wesentlicher Vorteil ist damit die Möglichkeit eröffnet, bei inhaltlichen Auswertungen gegebenenfalls nur solche Messungen zu berücksichtigen, die die Güteanforderungen zufriedenstellend erfüllen. Außerdem ist es erst unter diesen Bedingungen aussichtsreich, Umstände zu untersuchen, die für eine "erfolgreiche" Messung ausschlaggebend sind, um zukünftige Befragungen dann so einzurichten, daß nachteilige Kontexteffekte minimiert werden.

Dieser letzte Gesichtspunkt ist für die Magnitude-Messung in Umfragen ganz besonders wichtig. Die Methode ist ja das Produkt psychophysischer Laboratoriumsarbeit, bei der Kontextbedingungen sich noch relativ leicht kontrollieren lassen. Ihre Anwendung im Interview aber ist Variationen der Erhebungssituation voll ausgesetzt. Soll es nicht zu einer Kluft zwischen dem Anspruch der Methode und der Praxis ihrer Anwendung kommen, müssen die Auswirkungen typischer Situationsfaktoren auf die Meßresultate daher bekannt und ausschaltbar sein. Am Kriterium der individuellen Skalierungsgüte, das die Magnitude-Messung bereitstellt, ist die Erforschung solcher Kontexteffekte prinzipiell möglich. Allerdings liegen entsprechende Untersuchungen bislang nicht vor.

Als Hilfestellung für potentielle Anwender wird deswegen im folgenden über die erste Untersuchung berichtet, mit der versucht wurde, die Kontextproblematik von Magnitude-Skalen auf breiter Basis anzugehen. Im Rahmen des ZUMABUS III 1979 wurden repräsentativ für die Bevölkerung der Bundesrepublik und West-Berlins 2020 Personen nach dem "Ansehen" befragt, das ihrer Meinung nach die Inhaber von 16 Berufen bei uns genießen. Die Befragten sollten ihre diesbezüglichen Antworten mit mehrmodalen Magnitude-Urteilen ausdrücken, und zwar mit proportionalen Zahlen (magnitude estimation) und unterschiedlich langen, zu zeichnenden Linien. Zu Vergleichszwecken wurde das Ansehen der Berufe auch auf konventionelle Art mit Hilfe kategorischer Antwortvorgaben skaliert. Das inhaltliche Ziel dieser bundesweiten Erhebung bestand darin, Vorarbeiten für die Konstruktion einer deutschen Berufsprestige-Skala zu leisten, wobei die Überzeugung bestimmend ist, daß eine Magnitude-Prestigeskala die Mängel herkömmlicher Prestigemessungen ausschalten und zu inhaltlich neuen Ergebnissen führen kann. Unter methodischen Gesichtspunkten stand demgegenüber die Absicht im Vordergrund, Merkmale der Befragten, der Interviewer und Prozeßfaktoren der Interviewsituation mit der erreichten individuellen Skalierungsgüte in Beziehung zu setzen, um die Wirkungen dieser Kontextbedingungen bestimmen zu können. Die nachstehenden Ausführungen beschäftigen sich ausschließlich mit den Ergebnissen dieser zweiten, methodischen Zielsetzung.

Zunächst (in Abschnitt 1) wird das Vorgehen und die Anlage der Untersu-

---

chung beschrieben und es werden wesentliche Teile des Interviewmaterials abgedruckt, um potentiellen Anwendern als Vorlage zu dienen. Abschnitt 2 enthält Ergebnisse in Bezug auf die Befragten- und Interviewerleistungen gemessen am Kriterium der individuellen Skalierungsgüte. Der 3. Abschnitt untersucht Kontexteffekte, d.h. den Einfluß, den Befragten- und Interviewermerkmale und die Interviewsituation auf die Qualität der Messungen haben.

## 1. Die Methode

### 1.1 Ein Blick in die Theorie

Ausschlaggebend für die Magnitude-Messung von Einstellungsintensitäten ist das ursprünglich von S.S. Stevens (1953) proklamierte psychophysische Potenzgesetz. Es besagt, daß in Bezug auf variable Reizintensitäten  $S_i$  (in Energiebeträgen) die korrespondierenden Empfindungsintensitäten  $\Psi_i$  als Potenzfunktion dieser Reizintensitäten variieren, so daß  $\Psi_i = S_i^\beta$ . Wenn  $R_i$  numerische Reaktionen im Sinne impliziter Verhältnisschätzungen sind und unterstellt wird, daß diese zur Empfindungsstärke proportional sind, gilt entsprechend  $R_i = a S_i^\beta$ . Die Potenzfunktion beschreibt eine Situation, in der einem geometrischen Zuwachs auf der S-Achse ein geometrischer Zuwachs auf der R-Achse entspricht, so daß als Kurzformel gilt: "gleiche Stimulusverhältnisse produzieren gleiche Empfindungsverhältnisse". Empirisch hat sich erwiesen, daß der Exponent  $\beta$  der psychophysischen Funktion für unterschiedliche Reizmodalitäten einen unterschiedlichen Wert annimmt, z.B. 1/3 bei der Wahrnehmung von Helligkeitsreizen, 2/3 bei Lautheit, 1/1 bei räumlicher Länge, 3/2 bei Gewicht usw.

Aus diesen modalitäts-spezifischen Eigenschaften unserer Wahrnehmung läßt sich ein für die Einstellungsmessung bedeutsamer Tatbestand ableiten. Wenn  $R_{a_i} = a A_i^\alpha$  und  $R_{b_i} = b B_i^\beta$  als zwei bewährte psychophysische Funktionen in Bezug auf die Reizmodalitäten A und B gegeben sind, ergibt sich für  $R_{a_i} = R_{b_i}$

$$A_i = (b/a)^{1/\alpha} B_i^{\beta/\alpha}.$$

Der Fall  $R_{a_i} = R_{b_i}$  darf dann angenommen werden, wenn Befragte in Reaktion

auf eine Stimulusserie  $C_i$  aufgefordert werden, ihre entsprechenden Empfindungsstärken mit Hilfe der beiden Modalitäten A und B auszudrücken. Jede für den Befragten manipulierbare und extensive Größe kommt hierfür in Betracht, neben dem Zahlenkontinuum etwa zu zeichnende Linien, deren Längen die relativen Intensitäten von Empfindungen markieren sollen, oder auch einstellbare Lautstärken, Helligkeiten, ausgeübter Handdruck usw. Es ist dann vorhersagbar, daß die beiden Reaktionsvektoren der benutzten Modalitäten eine Potenzbeziehung eingehen, für die der Exponent  $\beta/\alpha$  charakteristisch ist. Für diese Schlußweise ist es nicht erforderlich, daß die Stimuli aus C unabhängig von den subjektiven Einschätzungen metrisierbar sind, d.h. auch Einstellungsitems kommen hierfür in Frage. Vorauszusetzen ist nur, daß die Reaktionsmodalitäten A und B reagierenden Exponenten  $\alpha$  und  $\beta$  bekannt sind. Wenn bei einer empirischen Durchführung die beiden Reaktionsmodalitäten in einer Potenzbeziehung stehen, die sich durch einen Exponenten beschreiben läßt, der sich nicht signifikant von dem prognostizierten Wert  $\beta/\alpha$  unterscheidet, erfüllt die Messung das Gütekriterium der Magnitude-Messung. (Im Gegensatz zu dem gewohnten statistischen Sprachgebrauch sind also "nicht signifikante" Messungen brauchbare Messungen). Nach der psychophysischen Theorie ist dann sichergestellt, daß die Reaktionsskalen von A und B beide in gleicher Weise subjektive Intensitäten ausdrücken, die der Empfindungsintensität proportional sind und die deren verhältnissbewahrenden Charakter haben.

Für die praktische Arbeit in Umfragen haben sich als einfache Reaktionsmodalitäten Zahlen und Linien bewährt: Bei Zahlen-Antworten wird der Befragte aufgefordert, den sukzessiv dargebotenen Stimuli i, j, k, jeweils Zahlen in der Weise zuzuordnen, daß das Verhältnis der Zahlen dem Verhältnis seiner dazugehörigen Empfindungsstärken entspricht. Löst j bei einem Befragten eine z.B. doppelt so starke Empfindung als i aus, dann wird erwartet, daß der Befragte den numerischen Betrag, den er i zugewiesen hat, bei der Einschätzung von j verdoppelt, und für k usw. entsprechend. Bei der Wahl der ersten Zahl (für i) ist der Befragte frei; jede positive von Null verschiedene Zahl kommt hierfür in Betracht. Die nachfolgenden Stimuli hingegen müssen mit Zahlen versehen werden, die auf die erste Nennung derart bezogen sind, daß damit die relativen Verhältnisse zum Ausdruck kommen, in denen die Empfindungsstärken zueinander stehen. Beim Linienzeichnen ist prinzipiell genauso vorzugehen, nur daß

es hier darauf ankommt, daß die Längen der Linien, die der Befragte den Stimuli zuordnet, seine impliziten Verhältnisschätzungen wiedergeben.

Für beide Modalitäten ist typisch, daß ihr psychophysischer Exponent 1 ist, d.h.  $\beta/\alpha = 1$ . Die Bewertung der Skalierungsgüte bei einer Messung, die Zahlen und Linien benutzt, richtet sich also danach, ob zwischen der Reihe der genannten Zahlen und der Reihe der gezeichneten Linienlängen eine Potenzbeziehung mit der Steigung 1 nachweisbar ist oder nicht. Im positiven Fall ist es dann zulässig, Skalenwerte als geometrische Mittelung aus den beiden Modalitäten zu berechnen, d.h. der Skalenwert für Stimulus  $i$  ergibt sich als  $(A_i^\alpha B_i^\beta)^{1/2}$ .

## 1.2 Fragen und Antworten

Die bei ZUMA entwickelte Abfrage- und Beantwortungsprozedur für die Magnitude-Messung in Umfragen besteht aus drei Schritten, die als gemeinsamer Block in ein ansonsten konventionelles Interview eingeschaltet werden können. In der ersten Phase wird dem Befragten das Verfahren erklärt, und unter Anleitung des Interviewers schätzt er zunächst - erst mit Zahlen, dann mit Linienlängen - die relativen Größen von dargebotenen Kreisflächen ein. In der sich anschließenden Übung soll der Befragte selbständig - erst mit Linien und dann mit Zahlen - die "Verwerflichkeit" einer Reihe von "Vergehen und Verbrechen" beurteilen ("Unterschlagung von DM 20,--", "Fahrerflucht bei einem Unfall, bei dem jemand verletzt wird" usw.). Erst wenn diese beiden Übungsschritte erfolgreich absolviert sind, beginnt die eigentliche Skalierungsphase, in der der Befragte mit Linien und Zahlen seine Einstellungsintensitäten zu den interessierenden Fragen ausdrücken soll. Während aller drei Phasen antwortet der Befragte schriftlich. Er erhält dazu ein Antwortheft, in dem die unterschiedlichen Stimuli in einem Format, das ihm genügend Platz insbesondere für die Antworten mit Linien gibt, zusammen mit einem u.U. verkürzten Fragetext aufgeführt sind. Die vollständigen Fragen werden vom Interviewer jeweils mündlich formuliert. Bei der Einführung in die Methode ist der Interviewer gehalten, die Instruktionen nach Möglichkeit frei zu geben. Sicherheitshalber enthält der Fragebogen aber einen genauen Wortlaut, an dem er sich bei Bedarf orientieren kann.

Bei der Magnitude-Skalierung im ZUMABUS III wurde nach diesem Standardverfahren vorgegangen. Auf den folgenden Seiten sind die Vorlage für den Instruktionstext der beiden Übungsschritte (S. 11 bis S. 13) und der eigentliche Fragetext zur Messung des Berufsprestiges (S. 14) wiedergegeben. Eine Seite aus dem Antwortheft ist auf S. 15 reproduziert (Originalformat DIN A 4 quer). Den Interviewern wird getrennt vom Fragebogen das farbige "Merkblatt" mitgegeben, das auf S. 16 abgebildet ist. Die dort aufgeführten sechs Merksätze sollen die Schulung unterstützen, und an entsprechenden Stellen des Instruktions- und Fragetextes wird jeweils ausdrücklich auf bestimmte Merksätze hingewiesen. Auf S. 17 schließlich findet sich die Frageformulierung für die kategorische Skalierung des Berufsprestiges, die im ZUMABUS unmittelbar nach der Magnitude-Aufgabe vorgenommen wurde. Der wiedergegebene Text dieses Interviewteils enthält auch die Liste der 16 Stimulusberufe, die einzuschätzen waren. Der gesamte Block "Berufsprestige" war etwa in der Mitte des einstündigen Interviews plazierte.

Diese Erhebungsformate für Magnitude-Messungen sind das Ergebnis umfangreicher Voruntersuchungen. Dies gilt insbesondere für die Formatierung der Antwortheftseiten für die Linienreaktionen (maximal 4 Stimuli pro Seite; DIN A 4 - Querformat; optische Länge der Stimulustexte; vgl. S. 15) und die entsprechenden Anweisungen an den Befragten (vgl. Merksätze IV - VI). Auch hier macht sich als Vorteil bemerkbar, daß die Magnitude-Messung über ein Gütekriterium verfügt: In der Entwicklung war es auf diese Weise möglich, alternative Erhebungsformate gegeneinander auszutesten und sich am Ende für die Form zu entscheiden, mit der das Gütekriterium empirisch am häufigsten erfüllt wurde.

In diesen praktischen Zusammenhang gehört schließlich auch die Benennung, die für das Magnitude-Verfahren sowohl den Befragten als auch den Interviewern gegenüber gewählt wird. Als ein gut verständlicher Name hat sich "Vergleichsantwort-Methode" erwiesen. In den ersten Magnitude-Studien, die ZUMA unternahm, wurden Interviewer und Befragte noch mit dem Terminus "Psychophysik" erschreckt.

(Fortsetzung des Textes auf Seite 18)



Nr.

40.

ACHTUNG INTERVIEWER !

EINFÜHRUNG IN DIE VERGLEICHSA NTWORT-METHODE

Anhand der folgenden vier Übungsbeispiele erklären Sie dem Befragten möglichst kurz, wie das Antwortheft auszufüllen ist.  
Sie können dies relativ frei, d.h. mit eigenen Worten tun.  
Die Formulierungen hier im Fragebogen sollen dabei Hilfestellung leisten.  
"Zeigen" ist besser als viele Worte!

Arbeiten Sie bitte unbedingt auch Ihre schriftlichen Interviewer-  
anweisungen durch.  
Die sechs Regeln auf Ihrem M e r k b l a t t (Liste X) sollten Ihnen stets ge-  
läufig sein. Bitte behalten Sie jeden Übungsschritt des Befragten und auch die  
einzelnen Antwortstufen im Auge, um notfalls korrigierend einzugreifen.

Wenn Sie selbst die (wenigen) wichtigen Regeln beherrschen, werden Sie  
ein guter "Lehrmeister" sein.

Wir kommen nun zu unserem letzten Thema, das nur eine Frage enthält. Sie sollen diese jedoch auf eine  
andere Weise beantworten als Sie es bisher getan haben. Vorhin haben Sie ja Ihre Meinung zu bestimmten  
Fragen mit Hilfe einer Skala ausgedrückt, auf der Sie einen bestimmten Punkt bzw. eine Stufe herausuchten.

Man kann nun sein ganz persönliches Urteil noch sehr viel genauer ausdrücken.

Wie das geht, möchte ich Ihnen anhand einiger Beispiele zeigen, die hier in diesem Antwortheft stehen, das  
Sie selbst ausfüllen werden.

Antwortheft und Bleistift dem Befragten übergeben!

## 1. ÜBUNG (Antwortheft Seite 1)

Schlagen Sie jetzt bitte die erste Seite auf. Mit diesen verschiedenen großen Kreisen hier möchte ich Ihnen  
erklären, was Sie tun sollen.

Sie können sich vorstellen, daß man seine persönliche Meinung über die Größenunterschiede dieser fünf Kreise  
mit Zahlen ausdrücken kann: also daß man den größeren Kreisen höhere Zahlen und den kleineren entsprechend  
niedrigere Zahlen zuordnet.

Wenn man dem ersten Kreis hier (zeigen!) irgendeine beliebige Zahl gibt, - z.B. eine 10, wie groß ist  
dann ungefähr der zweite Kreis? (gleich fortfahren!) Gesetzt den Fall, der zweite Kreis würde auf Sie den Eindruck  
machen, daß er etwa doppelt so groß wie der erste Kreis ist; - was würden Sie dann für eine Zahl neben diesen  
zweiten Kreis schreiben, um seine Größe im Vergleich zum ersten auszudrücken?

Wenn Befragter ungefähr "20" sagt, Befragten anweisen, auf die gepunkteten  
Linien rechts vom ersten Kreis eine 10 und rechts vom zweiten  
Kreis eine 20 zu schreiben.

Und um wievielfach größer als der erste Kreis ist Ihrer Meinung nach der dritte Kreis ungefähr? - Und mit welcher  
Zahl würden Sie das ungefähr ausdrücken, wenn Sie dem ersten Kreis eine 10 gegeben haben?

Die Schätzung für den dritten Kreis sollte etwa bei 30 bis 40  
liegen; Befragter sollte seine Schätzung selbst eintragen!

Nr.

Und was ist mit dem vierten Kreis? Wie groß ist der ungefähr, wenn der erste Kreis eine Größe von 10 hat?

Falls Befragter nicht von selbst antwortet:

Nun, ich finde den vierten Kreis ungefähr halb so groß wie den ersten; und da wir dem ersten Kreis eine 10 gegeben haben, muß der kleine Kreis ungefähr eine Größe von 5 haben.

Auch für den untersten Kreis eine Zahl hinschreiben lassen.

Falls Befragter noch unsicher ist, zusätzliche Erläuterungen geben!

Sie sehen, daß es bei den Antworten darauf ankommt, daß Ihre Zahlen immer auf die Zahl für den ersten Kreis bezogen werden. Diese erste Zahl wollen wir daher Ihre persönliche VERGLEICHSAHLE nennen:

Es ist die von Ihnen ausgewählte Zahl, mit der Sie alle weiteren ANTWORTZAHLEN vergleichen sollen.

Dabei brauchen Sie nicht etwa lange zu rechnen, sondern nur Ihren ungefähren Eindruck von den Kreisen mit Zahlen auszudrücken.

WICHTIG ist, daß Sie natürlich mit jeder beliebigen Zahl beim ersten Kreis anfangen können. Hier in dem Beispiel hätten wir ja auch dem ersten Kreis anstatt 20 z. B. die Zahl 50 geben können.

Können Sie mir sagen, welche Zahl dann für den zweiten Kreis richtig wäre? ... Und welche Zahlen müßten Sie dann den anderen Kreisen geben?

Befragten diese Zahlen in Klammern eintragen lassen!

Falls Befragter Verständnisschwierigkeiten hat, das Bisherige Schritt für Schritt ausführlich wiederholen!

## 2. ÜBUNG (Antwort auf Seite 2)

Blättern Sie jetzt bitte um. Man kann seine Meinung über die Größenunterschiede der Kreise ja auch mit verschiedenen langen Linien, die man neben die Kreise zeichnet, ausdrücken: Je länger eine Linie ist, die Sie zeichnen, umso größer finden Sie den entsprechenden Kreis.

Wir fangen damit an, indem wir eine Linie für den ersten Kreis hier daneben zeichnen (hinzeigen!).

Dies ist die VERGLEICHSLINIE für die anderen Kreise. - Es ist nicht schlimm, wenn die Linien nicht ganz gerade sind.

Gesetzt den Fall, der zweite Kreis würde auf Sie wieder den Eindruck machen, daß er etwa doppelt so groß wie der erste ist - wie lang müßte dann Ihre Linie für den zweiten Kreis ungefähr sein? Zeichnen Sie sie bitte einmal ein.

Wenn Befragter eine etwa doppelt so lange Linie zeichnet, zum dritten Kreis weitergehen und eine Linie zeichnen lassen, usw.

Überprüfen, ob Befragter das Prinzip verstanden hat, sonst immer nachfragen: "Um wievielfach größer oder kleiner ist der Kreis im Vergleich zum ersten?" ...)

Sie sehen also im Prinzip geht das genauso wie mit den Zahlen: Man bestimmt zunächst eine VERGLEICHSLINIE und zeichnet die ANTWORTLINIEN immer im Vergleich zu dieser ersten Linie für den ersten Kreis. Dabei ist es ganz gleichgültig, wie lang Ihre VERGLEICHSLINIE ist; Sie müssen sie nur immer bei den anderen Einschätzungen im Auge behalten!

Nr.

Eines sollten Sie noch wissen: Falls der Platz für eine Linie, die Sie zeichnen wollen, nicht ausreicht, wenn Sie am rechten Papierrand angekommen sind, fangen Sie einfach wieder von vorne an und setzen die Linie hier drunter beliebig oft fort (hinzeigen!).

Merksätze II und IV (Liste X) beachten!  
Nur wenn nötig, entsprechende Erklärungen geben!

### 3. ÜBUNG (Antwortheft Seite 3)

Damit Sie das Antworten mit Linien und Zahlen jetzt einmal an einer richtigen Frage ausprobieren können, haben wir auf der nächsten Seite noch ein kurzes Beispiel. Sie sehen hier einige Vergehen und Verbrechen aufgeführt. Wir möchten wissen, wie verwerflich oder schlimm Sie persönlich diese Vergehen und Verbrechen finden.

Bitte beantworten Sie diese Frage für die einzelnen Vergehen und Verbrechen, indem Sie zunächst Ihre Meinung mit den Linien ausdrücken. Gehen Sie bitte genauso wie eben bei den Kreisen vor: Zeichnen Sie also zunächst eine beliebig lange VERGLEICHSLINIE für das erste Vergehen (Ladendiebstahl eines Schmuckstückes im Wert von DM 10,-) und zeichnen Sie dann für alle anderen Vergehen und Verbrechen auf dieser Seite ANTWORTLINIEN, die Ihre Meinung über die Schwere dieser Delikte im Vergleich dazu ausdrücken.

Bevor Sie anfangen, lesen Sie sich die ganze Seite kurz einmal durch. Es ist dann einfacher für Sie, eine VERGLEICHSLINIE festzulegen; Sie können die VERGLEICHSLINIE zwar beliebig lang machen, sie sollte aber nicht zu kurz, aber auch nicht zu lang sein, damit Sie die anderen Linien dann noch bequem zeichnen können.

Merksätze II und IV (Liste X) beachten!  
Nur wenn nötig, entsprechende Erklärungen geben!

### 4. ÜBUNG (Antwortheft Seite 4)

Auf der nächsten Seite sind die Vergehen und Verbrechen noch einmal aufgeführt. Drücken Sie jetzt bitte Ihre Meinung über die Schwere der Vergehen und Verbrechen auch noch mit Zahlen aus.

Sie beginnen natürlich wieder mit der Wahl einer beliebigen VERGLEICHSSZAHL, auf die Sie dann alle Ihre ANTWORTZAHLEN beziehen.

Merksätze II und V (Liste X) beachten!  
Nur wenn nötig, entsprechende Erklärungen geben!

79/6231

Nr.

41 A (Antwortheft Seite 5 - 8)

Blättern Sie jetzt bitte um zu unserer eigentlichen Frage. Auf den nächsten Seiten stehen insgesamt 16 bekannte Berufe. Ich möchte jetzt gerne wissen:

- welches A n s e h e n diese 16 Berufe heute Ihrer Meinung nach haben;  
d. h., wie sehr Leute mit diesen Berufen in unserer Gesellschaft heute  
g e a c h t e t werden.

Wenn Sie diese Frage jetzt zunächst mit Linien beantworten, denken Sie bitte daran, daß Sie als erstes eine VERGLEICHSLINIE brauchen. Lesen Sie aber vorher alle 16 Berufe kurz durch (bis zum bunten Trennblatt!). Sie können dann Ihre VERGLEICHSLINIE besser festlegen.

Zeichnen Sie nun zu jedem Beruf eine ANTWORTLINIE, die sich auf Ihre gewählte VERGLEICHSLINIE für den ersten Beruf - den Elektroinstallateur - bezieht. Unter Umständen müssen Sie ab und zu einmal zu dieser VERGLEICHSLINIE zurückblättern.

Kontrollieren Sie unauffällig, ob der Befragte richtig vorgeht!  
Merkblatt, vor allem die Punkte II, IV und V beachten!  
Notfalls erklärend eingreifen!

41 B Antwortheft Seiten 10 - 13

Wenn Sie jetzt damit fertig sind, blättern Sie bitte das farbige Trennblatt um und beantworten Sie dieselbe Frage noch einmal - nur jetzt mit Zahlen.

Wählen Sie sich also bitte zuerst wieder eine VERGLEICHSZAHLE für den Elektroinstallateur und beziehen Sie alle Ihre ANTWORTZAHLEN darauf.

Welches Ansehen haben diese Berufe heute  
Ihrer Ansicht nach?

Elektroinstallateur

VERGLEICHSLINIE

Studienrat

ANTWORTLINIE

Fabrikarbeiterin

ANTWORTLINIE

Polizist

ANTWORTLINIE

## MERKSÄTZE FÜR DIE VERGLEICHSANTWORT-METHODE

- I. Erst alle Fragepunkte einmal durchsehen.
- II. Eine VERGLEICHSLINIE bzw. VERGLEICHSAHL wählen, die in Hinsicht auf die nachfolgenden Fragepunkte "bequem" ist.
- III. Alle weiteren Fragepunkte im Vergleich zu dieser VERGLEICHSLINIE bzw. VERGLEICHSAHL einschätzen. Dabei die Frage stellen: "Um wievielfach größer oder kleiner sollen meine ANWORTLINIEN bzw. -ZAHLEN im Vergleich zum ersten Fragepunkt sein?"
- IV. Bei den Linien beachten:
  - Mit den Linien immer auf der linken Papierhälfte etwa unter der Schrift beginnen.
  - Wenn das Papier für eine Linie nicht ausreicht, die Linie am rechten Papierrand abbrechen und links unter der abgebrochenen Linie fortsetzen - sooft Sie wollen.
  - Alle Linien durchgehend zeichnen und nicht mehrere kleinere Linienstücke untereinander.
- V. Bei Zahlen beachten:
  - Keine Zahl darf die Größe 0 haben.
  - Negative oder Minuszahlen sind ebenfalls nicht erlaubt.
- VI. Wenn Sie einen Fragepunkt nicht einschätzen können, schreiben Sie ein Fragezeichen (" ? ") an die Stelle der ANWORTLINIE bzw. -ZAHLE; das entspricht "Weiß nicht / Keine Meinung".

Nr.

42. Legen Sie das Heft nun bitte zur Seite.

Zum Schluß möchte ich Sie bitten, die Frage noch einmal auf eine andere Weise zu beantworten, und zwar mit Hilfe dieser "Leiter" hier.

Skala 5 vorlegen!

Sagen Sie mir bitte zu jedem Beruf, den ich Ihnen jetzt nenne, ob er Ihrer Meinung nach ein eher hohes oder ein eher geringes Ansehen in unserer Gesellschaft heute besitzt.

Auf der Leiter bedeutet "9" ein sehr hohes Ansehen und "1" ein sehr geringes Ansehen. Sie können jetzt mit den Zahlen von 1 bis 9 ausdrücken, wie groß das Ansehen eines Berufes ist. Sie können die Berufe auf jeder dieser Stufen einordnen, aber nicht d a z w i s c h e n. Sagen Sie mir zu jedem Beruf nur die Nummer der betreffenden Stufe.

Berufe bitte einzeln vorlesen!

Skalenwert

	1	2	3	4	5	6	7	8	9
Elektroinstallateur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Studienrat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fabrikarbeiterin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Polizist	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bauingenieur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Volksschullehrerin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Taxiunternehmer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Maschinenschlosser	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Maurer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sekretärin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Arzt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kraftfahrzeugmechaniker	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Briefbote	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Krankenschwester	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bauarbeiter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Industrie Kaufmann	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## 1.3 Interviewerschulung

Solange die Magnitude-Technik nicht zum Standardrepertoire der Umfrageforschung gehört, stellt die Schulung der Interviewer ein nicht zu unterschätzendes Problem dar. Dieses Problem ist besonders dann ernst zu nehmen, wenn die Feldarbeiten ausgelagert und an ein kommerzielles Institut delegiert werden sollen. Zumindest im Fall repräsentativer Erhebungen scheidet die Möglichkeit mündlicher Schulungen dann aus Kostengründen aus. Eine spezielle Schulung der Interviewer ist aber notwendig, weil das Magnitude-Verfahren an die Interviewer ganz neuartige Anforderungen stellt. Es setzt eine gewisse Umdefinition des eingefahrenen Rollenselbstverständnisses der Interviewer voraus, insofern sie den Befragten ja zunächst etwas "beibringen" müssen und damit aus der relativen Passivität, die von ihnen sonst in der Interviewsituation gefordert wird, heraustreten müssen.

Um zu überprüfen, welche Auswirkungen der Verzicht auf eine mündliche Schulung für Magnitude-Messungen hat, wurde für den ZUMABUS III ein gestaffeltes Trainingsverfahren gewählt, bei dem außerdem unterschiedliche Schulungsbedingungen realisiert wurden. Es wurde wie folgt vorgegangen:

1. Zunächst wurde eine "Einführung in die Vergleichsantwort-Methode" entworfen, die zugleich ein "Übungsheft für Interviewer" darstellte. Die sukzessiven Lernschritte des Heftes verlangten von den Interviewern die Beantwortung von Übungsaufgaben und selbstdurchzuführende Skalierungen. Das Übungsheft kam im Rahmen einer Voruntersuchung unter Mitwirkung von Interviewern ZUMAs und des kooperierenden Instituts (INFRATEST) zustande. Ohne theoretisch weitschweifig zu sein, wurde versucht, den Interviewern den Sinn der Methode in Gegenüberstellung zu ihnen vertrauten Verfahren nahezubringen, und das Vorgehen an Beispielen zu demonstrieren. Am Ende des Übungsheftes sollten die Interviewer den Fragenkomplex "Berufsprestige" selber beantworten.

2. Etwa sechs Wochen vor Beginn der Feldphase wurde die Endversion der Übungshefte an ca. 400 Interviewer verschickt. Die Hefte waren auszufüllen und zurückzusenden.

3. Parallel dazu wurde eine Gruppe von Interviewern aus dem Einzugsgebiet



ZUMAs zu einer mündlichen Schulung eingeladen und ausführlich in das Magnitude-Verfahren eingewiesen. Die mündlich geschulten Interviewer mußten die Übungshefte anschließend ebenfalls bearbeiten und einsenden.

4. Daraufhin wurden alle zurückgesandten Übungshefte ausgewertet. Das Ziel dabei war, nur jene Interviewer für den Einsatz bei der Feldarbeit auszuwählen, die die Methode sicher beherrschten. Das Auswahlkriterium war in erster Linie das Gütekriterium der Magnitude-Messung: Nur Interviewer, deren Skalierungen einen nicht signifikant von 1 verschiedenen Exponenten zwischen Linien- und Zahlen-Antworten aufwiesen, waren zuzulassen. Zusätzlich wurden aber auch die schriftlichen Kommentare, die von den Interviewern verlangt worden waren, berücksichtigt. Auf diese Weise sollte u.a. vermieden werden, daß Interviewer mit starken Vorbehalten gegen die neue Methode gezwungen wurden, damit zu arbeiten.

Die Auswertung der Übungshefte führte zur Unterscheidung "akzeptierter" und "nicht akzeptierter" Interviewer. Im Fortgang erwies es sich aber als nicht durchführbar, für die Feldarbeit am ZUMABUS nur akzeptierte Interviewer einzusetzen. Um alle sample points ausreichend besetzen zu können, war es unvermeidlich, auch auf nicht akzeptierte Interviewer zurückzugreifen. In dem Fall wurde versucht, die betreffenden Interviewer telephonisch zu kontaktieren und auf diesem Weg eine Nachschulung vorzunehmen. Das gelang nicht in jedem Fall. Darüber hinaus mußten sogar zusätzlich Interviewer eingesetzt werden, die weder schriftlich noch mündlich geschult waren. Zur Abrundung des designs war eine solche non-treatment Gruppe allerdings willkommen. Schließlich nahmen an der Untersuchung auch noch Interviewer teil, die zwar ein Übungsheft ausfüllten, aber verspätet zurückschickten, so daß unter ihnen nicht mehr rechtzeitig eine Trennung nach "akzeptiert" und "nicht akzeptiert" vorgenommen werden konnte. Diese Interviewer stellen eine unspezifische Mischgruppe dar. Ihre Notwendigkeit macht die Grenzen kontrollierter Forschung in diesem Bereich deutlich.

Insgesamt ergaben sich damit sechs Interviewergruppen unterschiedlicher Schulungsbedingungen, die in der Feldphase eingesetzt werden konnten. Tab. 1 beschreibt ihre Besetzungstärken.

Tab. 1

1. Schriftliche Schulung, (a) akzeptiert ....	182 (55.7%)
2. Schriftliche Schulung, (b) nicht akzeptiert .....	13 ( 4.0%)
3. Schriftliche Schulung, (c) nachgeschult ..	18 ( 5.5%)
4. Schriftliche Schulung, (d) unspezifiziert.	64 (19.6%)
5. Mündliche Schulung (alle akzeptiert) .....	18 ( 5.5%)
6. Ohne Schulung .....	32 ( 9.8%)

-----  
327

5. Mit den Fragebogenunterlagen erhielten alle eingesetzten Interviewer eine gleichlautende ausführliche Intervieweranleitung mit besonders gekennzeichnetem "Vergleichsantwort-Teil". Darin wurde auf Bedenken und Anregungen, die die Interviewer schriftlich oder mündlich vorgebracht hatten, eingegangen.

## 2. Interviewer- und Befragtenleistungen

Die Theorie der Magnitude-Messung mit ihrem Kriterium für Skalenqualität erlaubt es, faktisch vorgenommene Messungen unter Leistungsgesichtspunkten zu betrachten. Im vorliegenden Fall erstreckt sich diese Betrachtungsweise nicht nur auf die Befragten, sondern auch auf die beteiligten Interviewer. Wie gut beherrschen beide die Methode und wie hoch ist die Qualität der Resultate? Wie bei konventionellen Skalierungsverfahren richtet sich hier der Blick als erstes auf die Ausfälle. Während aber bei konventionellen Verfahren die Fälle ohne Messung paradoxerweise das einzige stichhaltige Kriterium für die Güte dieser Verfahren sind, kann die Magnitude-Messung ein über die Minimierung von Ausfällen hinausgehendes Ziel anstreben: Grundlage für die Bewertung ist auch, ob der Skalierungsexponent signifikant vom erwarteten Wert abweicht oder nicht. In Bezug auf beide Gesichtspunkte haben die ZUMABUS-Daten die folgenden Ergebnisse gebracht.

## 2.1 Ausfälle

Die Ausfallsrate läßt sich sinnvoll nur in Bezug auf die Befragten bestimmen. Mit Ausnahme der Interviewergruppe "ohne Schulung" liegt für alle eingesetzten Interviewer ja ein vollständig ausgefülltes Übungsheft mit der Prestigeskalierung vor. Aus Tab. 2 geht hervor, daß von den 2020 Befragten 224 kein Antwortheft ausgefüllt haben. Das sind 11.1%. Offenbar ist dieser Prozentsatz von den verschiedenen Schulungsbedingungen der Interviewer unabhängig. In Tab. 2 sind die entsprechenden Spaltenprozente eingetragen.

Tab. 2

	AKZEP- TIERT	NICHT AKZEPT	NACHGE SCHULT	MÖNDL GESCH	UNGE- SCHUL	NICHT SPEZIF	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW-HEFT	89.3	88.3	87.7	88.2	90.4	89.1	88.9	1796
KEIN HEFT	11.0	11.7	12.3	11.8	9.6	10.9	11.1	224
GESAMTPROZ.	58.5	3.8	5.6	5.9	8.3	17.9		
ABSOL.ZAHL.	1181	77	114	119	167	362		2020
Spaltenprozente								

Die Anzahl der Ausfälle geht - wie wir noch sehen werden - zum Teil auf das Konto der Interviewer; bestimmte Eigenschaften der Interviewer scheinen bei den Befragten die Tendenz zur Verweigerung der Magnitude-Aufgabe zu begünstigen. Unabhängig davon hat sich aber ein kleiner Teil der Interviewer überhaupt geweigert, mit der Methode zu arbeiten. Das betrifft neun Interviewer, von denen die Mehrzahl ungeschult in die Untersuchung geschickt wurde. Läßt man die Interviews dieser Interviewer bei der Berechnung der Ausfälle unberücksichtigt, verringert sich der Prozentsatz nicht zustandegekommener Magnitude-Skalierungen auf 8.6.

Zur Beurteilung dieser Anzahl der Ausfälle ist ein Vergleich mit der Ausfallhäufigkeit bei der kategorischen Skalierung des Berufsprestiges aufschlußreich. Die Ausfälle in der kategorischen Skalierung belaufen sich auf knapp 4%. Dieser Prozentsatz ergibt sich, wenn man die missing data über alle möglichen Nennungen der kategorischen Messung aufaddiert. Als Faustregel kann man daher formulieren: Bei einer herkömmlichen Kategoriale Skalierung haben wir (bei der vorliegenden Thematik) mit etwa 4% feh-

lenden Werten zu rechnen, bei Magnitude-Skalen verdoppelt sich diese Ausfallsquote in etwa.

## 2.2 Skalierungsgüte

Die Qualität der zustandegekommenen Messungen geht aus Tab. 3 hervor (mit ZUMAPSYCH berechnet; vgl. WEGENER, 1978). 7.1% - 15.9% der Befragten weisen signifikant von 1.0 verschiedene Exponenten der Beziehungen zwischen den Zahlennennungen und den Linienlängen auf. Diese Spannweite ergibt sich bei Zugrundelegung unterschiedlicher Signifikanzniveaus für den anzuwendenden t-Test (mit N-2 Freiheitsgraden; vgl. das ZUMAPSYCH-Manual). Bei einem mittleren Signifikanzniveau erweisen sich etwa 10% der Messungen als signifikant, d.h. daß diese Messungen nicht dem psychophysischen Modell entsprechen. Für Auswertungen sind sie deswegen zwar nicht völlig unbrauchbar. Sie genügen aber nicht dem angestrebten Skalenniveau.

Tab. 3

t (df=14)	4.14	5.36	6.71
BEFR.N.S.	84.1	89.8	92.9
BEFR.SIGN.	15.9	10.2	7.1
(N = 1796)	100.0	100.0	100.0

Die Magnitude-Prestigeskalierung, die die Interviewer in der Vorphase der Untersuchung in ihren Übungsheften vornehmen mußten, erlaubt es, Aussagen über die Auswirkung der dort gezeigten Skalierungsgüte auf die Leistungen der Befragten zu machen. Aus den dargestellten Gründen mußten nämlich etwa 7% der Interviewer eingesetzt werden, obwohl diese Interviewer den Test in den Übungsheften nicht "bestanden" und signifikant von 1 verschiedene Skalierungsexponenten zeigten ( $t = 5.36$ ). Aus Tab. 4 ist ersichtlich, daß Interviewer, die im Übungsheft schlecht abschnitten, bei ihren Interviews 12% signifikante Skalierungen bei den Befragten "produzierten", während Interviewer mit nicht-signifikanten Leistungen im Übungsheft nur etwa 10% "schlechte" Skalierungen bei ihren Befragten hervorriefen. Diese Tendenz spricht dafür, daß das Ausmaß, in dem die Interviewer die Aufgabe selber beherrschen, offenbar einen Einfluß auf ihre Fähigkeit zur Vermittlung hat. Auffällig ist, daß die Interviewer ohne

irgendeine Schulung hier am besten abschneiden. Sie produzieren nur 9% signifikante Skalierungen bei den Befragten. Das ist ein Hinweis darauf,

Tab. 4

	BEFR. N.S.	BEFR. SIGN.	GESAMT PROZ.	ABSOL. ZAHLEN
INT.N.S.	90.1	9.9	84.2	1512
INT.SIGN.	88.0	12.0	6.5	117
OHNE SCHUL.	91.0	9.0	9.3	167
GESAMTPROZ.	89.8	10.2		
ABSOL.ZAHL.	1613	183		1796

Zeilenprozente

daß die Intervieweranleitung und der Fragetext im Fragebogen für sich bereits hinreichend informativ waren, um motivierte Interviewer in die Methode einzuführen

Zeitaufwand: Interviews, bei denen ein Antwortheft ausgefüllt wurde, dauerten im Mittel 58.5 Minuten. Ohne Antwortheft, d.h. bei Interviews, bei denen keine Magnitude-Messung zustandekam, ist die mittlere Interviewzeit 52.4 Minuten. Etwa sechs Minuten an zusätzlicher Interviewzeit wird also für die Magnitude-Skalierung und die Übungsschritte in Anspruch genommen. Aufschlußreich ist, daß kein Zeitunterschied besteht, wenn man die Interviewzeit von Befragten mit signifikanten und von solchen mit nicht signifikanten Resultaten im Magnitude-Teil vergleicht.

### 3. Kontexteffekte

Die Anlage der Untersuchung stellt eine Reihe von Attributen sowohl der Befragten als auch der Interviewer zur Verfügung, so daß Fragen nach ihrem Einfluß auf 1. die Anzahl der Verweigerungen und auf 2. die Güte der resultierenden Messungen (der Befragten) beantwortet werden können. Einen Überblick geben die Tabellen 5 und 6, in denen die Ergebnisse von multiplen Regressionsanalysen dargestellt sind. Beide abhängigen Variablen sind dichotom, und die unstandardisierten Koeffizienten sind erwartungsgemäß niedrig. Die relative Bedeutsamkeit der untersuchten unabhängigen Variablen ist deswegen in erster Linie an der Rangordnung der absoluten Koeffizienten abzulesen.

Tab. 5

Abhängige Variable: Verweigerung (=0) vs. keine Verweigerung (=1)

RANG	BETA	UNABHAENGIGE VARIABLEN
1	-.2382	Alter des Befragten
2	.1295	Anzahl im ZUMABUS durchgeführter Interviews
3	.0908	Schulbildung des Interviewers
4	-.0733	Geschlecht des Befragten (m = 1, w = 2)
5	-.0735	Berufsprestige des Interviewers
6	.0639	Schulbildung des Befragten
7	-.0503	Alter des Interviewers
8	-.0475	Einkommen des Befragten
9	-.0418	Geschlecht des Interviewers (m = 1, w = 2)
10	.0323	Prestige des erlernten Berufs des Interviewers
11	.0252	Berufserfahrung des Interviewers in Monaten

Tab. 6

Abhängige Variable: Befragtenskalierung ist signifikant (=0) vs. nicht signifikant (=1)

RANG	BETA	UNABHÄNGIGE VARIABLEN
1	.1909	Anzahl 1978 durchgeführter Interviews
2	.1013	Schulbildung des Interviewers
3	.0855	Geschlecht des Befragten (m = 1, w = 2)
4	-.0819	Geschlecht des Interviewers (m = 1, w = 2)
5	.0792	Berufserfahrung des Interviewers in Monaten
6	-.0651	Anzahl im ZUMABUS durchgeführter Interviews
7	-.0556	Alter des Interviewers
8	-.0484	Einkommen des Befragten
9	-.0481	Berufsprestige des Interviewers
10	.0472	Alter des Befragten
11	.0445	Prestige des erlernten Berufs des Interviewers
12	-.0400	Berufsprestige des Befragten

## 3.1 Alter

Als eine sehr wichtige Einflußgröße erweist sich das Alter, und zwar vor allem der Befragten. Entgegen den Befürchtungen, die ein Großteil der Interviewer vor Beginn der Studie äußerten, ist es nicht der Fall, daß ältere Befragte mit dem Verfahren nicht zurechtkommen. Es ergibt sich vielmehr ein differenzierteres Bild: Jüngere Befragte verweigern seltener, die Antworthefte überhaupt auszufüllen; von denjenigen Befragten jedoch, die erst einmal damit anfangen und sich auf die Methode einlassen, produzieren ältere Befragte deutlich bessere Ergebnisse. Die Hemmschwelle liegt also bei älteren Menschen offenbar höher, ihre Leistungen sind aber, sofern sie diese Schwelle überwinden, besser. Darüber hinaus spiegelt die hohe Ausfallsrate bei älteren Befragten sicherlich auch die entsprechende Erwartungshaltung der Interviewer wider.

Das mittlere Befragtenalter bei Verweigerungen beträgt 59.5 Jahre, bei ausgefüllten Antwortheften hingegen nur 45.6 ( $p < .0001$ ). Das umgekehrte Bild ergibt sich bei der Betrachtung der Skalierungsgüte. Hier ist das mittlere Alter derjenigen Befragten, die eine signifikante, d.h. "schlechte", Skalierung abliefern, 42.0 Jahre, während Befragte mit nicht signifikanten Ergebnissen im Mittel 46.3 Jahre alt sind ( $p < .001$ ). Die Tabellen 7 und 8 geben einen Eindruck von der Verteilung der Ausfälle beziehungsweise der Skalierungsqualität auf sechs Altersgruppen.

Tab. 7

	18-30	31-40	41-50	51-60	61-70	71 UND ÄLTER	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	96.1	93.5	93.5	89.7	82.4	70.2	88.9	1783
KEIN HEFT	3.9	6.5	6.5	10.3	17.6	29.8	11.1	222
GESAMTPROZ.	21.7	19.9	16.1	15.5	14.2	12.7		
ABSOL.ZAHL.	435	398	323	310	284	255		2005

Spaltenprozente

Auch das Alter der Interviewer ist für beide Kriteriumsvariablen von Bedeutung, wenn auch vergleichsweise in einem geringen Ausmaß. Für beide abhängigen Variablen ist die Tendenz nachweisbar, daß jüngere Interviewer zu besseren Resultaten bei ihren Befragten gelangen als ältere. Of-

fenbar können jüngere Interviewer die Methode besser vermitteln und die Befragten eher motivieren. Allerdings ist eine Wechselwirkung mit dem Alter der Befragten festzustellen: Die Bedeutung des Intervieweralters für

Tab. 8

	18-30	31-40	41-50	51-60	61-70	71 U.	GESAMT	ABSOL.
						AELTER	PROZ.	ZAHL
BEFR. N.S.	92.8	90.6	91.7	94.6	94.4	96.1	93.0	1658
BEFR. SIGN	7.2	9.4	8.3	5.4	5.6	3.9	7.0	125
GESAMTPROZ.	23.4	20.9	16.9	15.6	13.1	10.0		
ABSOL.ZAHL.	418	372	302	278	234	179		1783

Spaltenprozente

den Prozentsatz von Verweigerungen ist bei jüngeren Befragten (bis 50 Jahre) größer als bei Befragten, die älter als 50 Jahre sind ( $p < .05$  vs.  $p < .10$ ). D.h., jüngere Interviewer sind besonders erfolgreich bei einer Befragtenpopulation unter 50 Jahren, mit der sich ältere Interviewer hingegen schwertun.

## 3.2 Geschlecht

Ein bemerkenswerter Effekt ergibt sich in Hinsicht auf das Geschlecht sowohl der Interviewer als auch der Befragten. Grob gesprochen gilt, daß weibliche Befragte wesentlich günstigere Ergebnisse produzieren als männliche (Tab. 9). Allerdings verweigern männliche Befragte seltener als weibliche (Tab. 10). Als Interviewer andererseits sind die Männer unter beiden Gesichtspunkten im Vorteil: Bei ihnen finden sich 3% weniger Verweigerungen des Magnitude-Teils als bei weiblichen Interviewern, und in etwa dasselbe gilt für den Prozentsatz signifikanter Skalierungen.

Es besteht allerdings die Tendenz, daß ältere weibliche Interviewer bei jüngeren weiblichen Befragten zu besseren Ergebnissen kommen als ihre männlichen Kollegen bei derselben Population: Bei weiblichen Befragten, die jünger als 50 Jahre sind, rufen über 50-jährige Interviewerinnen fast 3% weniger signifikante Messungen hervor als männliche Interviewer derselben Altersklasse.



Tab. 9

	BEFR. MAENNL.	BEFR. WEIBL.	GESAMT PROZ.	ABSOL. ZAHLEN
BEFR.N.S.	82.7	85.4	84.1	1504
BEFR.SIGN.	17.3	14.6	15.9	284
GESAMTPROZ.	46.5	53.5		
ABSOL.ZAHL.	831	957		1788

Spaltenprozente

Tab. 10

	BEFR. MAENNL.	BEFR. WEIBL.	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	90.1	86.7	88.8	1767
KEIN HEFT	9.9	13.3	11.2	223
GESAMTPROZ.	61.4	38.6		
ABSOL.ZAHL.	1221	769		1990

Spaltenprozente

### 3.3 Statusvariablen

An zweiter bzw. dritter Stelle auf der Rangliste der relativen Effekte steht in Bezug auf beide Kriterienvariablen die Schulbildung der Interviewer (in Jahren). Interviewer mit einer eher hohen schulischen Ausbildung vermitteln im Interview die Methode besser an die Befragten, so daß bei diesen seltener signifikante Skalierungsexponenten zu finden sind (Tab. 11) und es weniger häufig zu Ausfällen kommt (Tab. 12). Der Einfluß der Schulbildung der Befragten ist demgegenüber gering. Bei dem Verweigerungskriterium erscheint diese Variable an sechster Stelle (vgl. Tab. 13), in Hinsicht auf die Skalierungsgüte hingegen liegt überhaupt keine nennenswerte Einflußnahme der Schulausbildung vor.

Tab. 11

	HAUPT- SCHULE INT.	MITTL. REIFE INT.	ABITUR FACHHR INT.	GESAMT PROZ.	ABSOL. ZAHLEN
BEFR. N.S.	81.9	83.3	86.8	83.8	1297
BEFR.SIGN.	18.1	16.7	13.2	16.2	251
GESAMTPROZ.	32.4	43.2	24.4		
ABSOL.ZAHL.	502	669	71		1548

Spaltenprozente

Tab. 12

	HAUPT- SCHULE INT.	MITTL. REIFE INT.	ABITUR FACHHR INT.	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	85.7	89.3	92.9	88.9	1548
KEIN HEFT	14.3	10.7	7.1	11.1	194
GESAMTPROZ.	33.6	43.0	23.4		
ABSOL.ZAHL.	586	749	407		1742

Spaltenprozente

Tab. 13

	HAUPT- SCHULE BEFR.	MITTL. REIFE BEFR.	ABITUR FACHHR BEFR.	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	86.2	92.3	94.6	88.3	1638
KEIN HEFT	13.8	7.8	5.4	11.7	216
GESAMTPROZ.	67.1	21.6	11.4		
ABSOL.ZAHL.	1244	400	210		1854

Spaltenprozente

Andere Statusvariablen sind vergleichsweise unbedeutend. Interessant ist allerdings ein Merkmal der Interviewer: Das Prestige ihres gegenwärtigen Berufs (in Treiman-Scores) weist einen geringen Effekt in Bezug auf die Skalierungsqualität der Befragten auf, jedoch in unerwarteter Richtung. Interviewer mit einem eher niedrigen Sozialprestige sind in der Vermitt-

lung der Methode besser. Ein umgekehrtes Ergebnis findet sich allerdings, wenn man nach dem Prestige des erlernten Berufs fragt. Hier wirken sich höhere Niveaus günstiger aus, und zwar sowohl in Hinsicht auf die erzeugten Verweigerungen als auch in Hinsicht auf die Befragtenleistungen. Für diesen Befund bietet sich eine plausible Erklärung an: Man kann davon ausgehen, daß professionelle Interviewer in den seltensten Fällen außerhalb dieser Tätigkeit einen befriedigenden und ihnen sozial angemessenen Lebensberuf gefunden haben. Sie wären anders nicht auf die Umfragearbeit angewiesen. Wahrscheinlich ist daher in vielen Fällen, daß der ursprünglich erlernte Beruf ihr Herkunftsniveau zutreffender kennzeichnet als die Angabe des heute ausgeübten Berufs, der oft einen Abstieg markieren dürfte. Es ist deswegen nicht verwunderlich, daß die Hierarchisierung der Ursprungsberufe positiv mit dem gezeigten Interviewererfolg korreliert, da angenommen werden kann, daß ein eher hoher sozialer Herkunftstatus sich auf die Arbeit als Interviewer günstig auswirkt.

### 3.4 Übungseffekte und Interviewererfahrung

Fällt Interviewern mit einer langen Berufserfahrung als Interviewer die Anwendung der neuen Methode leichter und erreichen sie bessere Resultate als "Neulinge"? Die Interviewererfahrung wurde differenziert erhoben, und zwar 1. durch die Angabe der Anzahl der im abgelaufenen Jahr (1978) ausgeführten Einzelinterviews und 2. durch die Angabe der Dauer der Tätigkeit als Interviewer in Monaten. Dabei wurde davon ausgegangen, daß sich die Intensität der Berufserfahrung als Interviewer nicht unbedingt mit der Dauer decken muß, die ein Interviewer schon als Interviewer arbeitet.

Während beide Variablen nur in einem unbedeutenden Sinne für die Verweigerungsrate verantwortlich zu sein scheinen, haben sie einen deutlichen Effekt auf die Skalierungsgüte, die die Befragten im Interview erreichen. Die Anzahl der im vergangenen Jahr ausgeführten Interviews ist regressions-analytisch sogar der wichtigste Faktor, und zwar in der zu erwartenden Richtung, daß mehr besser ist. Während Interviews, bei denen der Skalierungsexponent des Befragten sich als von 1 verschieden herausstellte, von Interviewern durchgeführt wurden, die 1978 im Mittel 30 Interviews machten, sind es bei den Interviewern mit nicht signifikanten

Ergebnissen fast 40 Interviews ( $p < .05$ ) Die reine Beschäftigungszeit als Interviewer wirkt in dieselbe Richtung, aber weniger ausgeprägt.

Bemerkenswert ist im übrigen, daß die Leistungen der Interviewer in ihren Übungsheften ebenfalls stark von der Anzahl der im vergangenen Jahr durchgeführten Interviews abhängt. Erfahrene Interviewer haben die Übungshefte erfolgreicher, d.h. mit einem größeren Prozentsatz nicht signifikanter Resultate ausgefüllt. Es ist anzunehmen, daß sich der tendenziell günstige Einfluß der Skalierungsqualität im Übungsheft auf die Befragtenleistungen (vgl. Tab. 4) von daher erklärt.

Man kann erwarten, daß die Interviewer in dem Maße, in dem sie die Vergleichsantwort-Methode wiederholt anwenden, im Umgang mit dem Verfahren Routine und Sicherheit gewinnen, und daß sich dies positiv auf die Befragtenleistungen auswirkt. Die Abschätzung solcher Trainingseffekte läßt sich über zwei Indikatoren bestimmen: 1. die Anzahl der Interviews, die ein Interviewer in der vorliegenden Studie durchgeführt hat, und 2. über die "Listen-Nummer", mit der die Interviewer die Reihenfolge ihrer realisierten Interviews in der Studie angeben. Beide Indikatoren sind allerdings für die Analyse von Trainingseffekten nur mit Vorbehalt zugrunde zu legen, weil sich im Nachhinein nicht kontrollieren läßt, ob mehrere Interviewer nicht unter ein und derselben "Sammelnummer" gearbeitet haben (eine Praxis, von der die Umfrageinstitute nur ungern abgehen), und ob die Sequenzen der Listen-Nummern tatsächlich die Reihenfolge der Interviews kennzeichnen.

Unter diesen Vorbehalten ist es erstaunlich, in welcher Deutlichkeit beide Variablen mit den Kriteriumsvariablen kovariieren. Die Anzahl der in der Studie pro Interviewer durchgeführten Interviews erscheint in Bezug auf die Verweigerungs-/Nicht-Verweigerungsrate bereits an zweiter Stelle der Beeinflussung. Wer viele Interviews durchführt, produziert weniger Verweigerungen im Magnitude-Teil. In der mittleren Anzahl ausgeführter Interviews unterscheiden sich Interviewer, die Ausfälle beziehungsweise keine Ausfälle erzeugen, auf dem .001-Signifikanzniveau (im Mittel 5.8 vs. 9 Interviews). Tab. 14 verdeutlicht die abnehmende Anzahl von Verweigerungen bei steigender Interviewzahl pro Interviewer.

Tab. 14

	1 IN- TERVIEW	2-5 IN- TERVIEWS	6-15 IN- TERVIEWS	16-20 IN- TERVIEWS	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	55.6	81.9	95.5	93.6	89.6	1719
KEIN HEFT	44.4	18.1	4.5	6.4	10.4	199
GESAMTPROZ.	2.8	34.0	57.5	5.7		
ABSOL.ZAHL.	54	652	1102	110		1918

Spaltenprozente

In Bezug auf die Skalierungsqualität bei den Befragten erscheint die Variable "Anzahl der jetzt durchgeführten Interviews" immerhin bereits an sechster Stelle der Effekte. Erstaunlicherweise dreht sich die Richtung des Einflusses hier jedoch um: Tendenziell sind die Interviewer die besseren Vermittler der Methode, die in der Untersuchung eher wenige Interviews durchführen. Dieser Befund wird im Mittelwertsvergleich zwar nicht signifikant, aufschlußreich ist aber eine Häufigkeitstabelle, die die Prozentsätze "signifikanter" Befragter pro Interviewergruppe mit unterschiedlicher Interviewanzahl aufführt (Tab. 15). Man sieht, daß die Interviewer, die überhaupt nur ein Interview durchgeführt haben, 10% "signifikante" Befragte produzieren; bei 2-5 Interviews sinkt dieser Prozentsatz, um bei mehr als 16 Interviews dann stark anzusteigen. Es scheint hier zu gelten, was als Regel für den Einsatz von Interviewern überhaupt beachtet werden muß: Wer in ein und derselben Untersuchung sehr viele Interviews macht, verliert im Laufe die Motivation und die Sensibilität für das Instrument, so daß seine Leistung nachläßt.

Tab. 15

	1 IN- TERVIEW	2-5 IN- TERVIEWS	6-15 IN- TERVIEWS	16-20 IN- TERVIEWS	GESAMT PROZ.	ABSOL. ZAHLEN
BEFR.N.S.	90.0	91.6	89.5	80.6	89.6	1541
BEFR.SIGN.	10.0	8.4	10.5	19.4	10.4	178
GESAMTPROZ.	1.7	31.1	61.2	6.0		
ABSOL.ZAHL.	30	534	1052	103		1719

Spaltenprozente

Ein von diesen Befunden abweichendes Bild ergibt sich bei Berücksichtigung der Listen-Nummern, durch die die Reihenfolge der ausgeführten Interviews rekonstruiert werden kann. Hier fällt auf, daß nicht nur der An-

teil "schlechter" Befragtenskalierungen (Tab. 16), sondern auch die Verweigerungsanzahl stetig mit der Sequenz der abgeleiteten Interviews zunimmt (Tab. 17). M.a.W., ein positiver Trainingseffekt ist nicht nachweisbar. Im Gegenteil, die Leistungen verschlechtern sich bei späteren Interviews. Aus diesem Resultat ist die Empfehlung abzuleiten, daß eine mittlere Anzahl von Interviews - vielleicht maximal 5 oder 6 - die optimale Zahl von Interviews ist, die der einzelne Interviewer durchführen sollte. Inwieweit eine solche Forderung realisierbar ist, ist eine ganz andere Frage.

Tab. 16

	1. IN- TERVIEW	2.-5. IN- TERVIEW	6.-16. IN- TERVIEW	GESAMT PROZ.	ABSOL. ZAHLEN
BEFR.N.S.	92.7	90.9	87.9	89.8	1605
BEFR.SIGN.	7.3	9.1	12.1	10.2	182
GESAMTPROZ.	9.2	38.0	52.9		
ABSOL.ZAHL.	164	679	944		1787

Spaltenprozenzte

Tab. 17

	1. IN- TERVIEW	2.-5. IN- TERVIEW	6.-16. IN- TERVIEW	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	90.1	89.8	87.7	88.9	1787
KEIN HEFT	9.9	10.2	12.3	11.1	223
GESAMTPROZ.	9.1	37.6	53.3		
ABSOL.ZAHL.	182	756	1072		2010

Spaltenprozenzte

### 3.5 Prozeßvariablen

Der Ablauf eines Interviews, insoweit er durch die Einschätzungen der Interviewer erfaßbar ist, hat einen deutlichen Einfluß auf die Skalierungsleistungen. Interviews, bei denen andere Personen als der Befragte anwesend waren, führen öfter zu signifikanten Skalierungsergebnissen als solche, die ungestört durchgeführt wurden. Bei Störung verdoppelt sich der Anteil signifikanter Skalierungsexponenten in etwa. Eingriffe in das Interview durch Drittpersonen verschlechtern das Resultat ebenfalls. Auffällig ist allerdings, daß nicht gestörte Interviews häufiger Verweige-

rungen des Magnitude-Teils erleben als Interviews, bei denen Dritte dabei waren (13% vs. 9% Verweigerungen;  $p < .005$ ). Offenbar ist ein Befragter im Beisein von ihm bekannten Personen aufgeschlossener und es fällt ihm leichter, die Aufgabe als kurzweilige Abwechslung zu begreifen.

Die von den Interviewern einzuschätzende Antwortbereitschaft der Befragten wirkt sich auf die beiden Kriteriumsvariablen in unterschiedlicher Weise aus. Es zeigt sich wider Erwarten, daß die Befragtenleistungen im Sinne der Exponentenprüfungen besser sind, wenn die Antwortbereitschaft der Befragten im Gesamtinterview als nur gering eingeschätzt werden. Vielleicht spricht dies dafür, daß gerade "antwortfaule" Probanden durch die Vergleichsantwort-Methode besonders motiviert werden. Andererseits korreliert allerdings die Bereitschaft der Befragten, überhaupt ein Antwortheft auszufüllen, mit ihrer eingeschätzten allgemeinen Antwortbereitschaft (Tab. 18). Die allgemeine Antwortbereitschaft von etwa 20% der Befragten wird von den Interviewern als "mittelmäßig" oder "schlecht" eingestuft, und diese Befragten weigern sich überproportional häufig, mit der Magnitude-Methode zu arbeiten. Offenbar müssen wir bei Umfragen mit einem festen Stamm notorischer Verweigerer rechnen, und auch das Vergleichsantwort-Verfahren scheint nicht in der Lage zu sein, diese Befragten in ihrer Haltung umzustimmen.

Tab. 18

	GUT	MITTEL	SCHLECHT	GESAMT PROZ.	ABSOL. ZAHLEN
ANTW.HEFT	93.3	80.5	61.1	89.8	1720
KEIN HEFT	6.7	19.5	38.9	10.2	195
GESAMTPROZ.	78.3	17.9	3.8		
ABSOL.ZAHL.	1500	343	72		1915

Spaltenprozente

#### 4. Schlußfolgerungen

Die Ergebnisse der skizzierten Methodenanalyse machen Zweierlei deutlich:

1. Magnitude-Messung ist als ein Instrument für die Umfrageforschung

auch in repräsentativen Untersuchungen anwendbar. Im allgemeinen liefern die Befragten die angestrebte Skalierungsqualität, allerdings liegt beim gegenwärtigen Stand der Methodenentwicklung die Ausfallquote etwa um das Doppelte höher als bei herkömmlichen Verfahren.

2. Es gibt offenbar "kritische" Befragtengruppen, bei denen die Methode unterdurchschnittlich effektiv ist. Ebenso unterscheiden sich auch die Interviewer hinsichtlich des Erfolgs, den sie bei ihren Befragten mit dem Verfahren erzielen. In beiden Fällen ist das Ausmaß dieser Effekte allerdings nicht derart gravierend, daß die Anwendbarkeit der Magnitude-Methode in repräsentativen Untersuchungen in Frage gestellt wäre oder daß das Verfahren bei bestimmten Befragten- oder Interviewergruppen überhaupt versagen würde.

Man muß sich in Bezug auf beide Resultate vergegenwärtigen, daß konventionelle Meßverfahren einer vergleichbaren Methodenuntersuchung nicht zugänglich sind, und daß in Bezug auf sie Aussagen über die Faktoren, die in der Anwendung die Skalierungsqualität beeinflussen, deswegen nicht gemacht werden können. Das Qualitätskriterium hingegen, über das die Magnitude-Messung verfügt, erlaubt es auf der Basis von Untersuchungen wie der vorliegenden, "kritische" Befragten- und Interviewergruppen (sowie Interaktionen zwischen beiden) zu isolieren und ihnen bei zukünftigen Untersuchungen eine besondere Aufmerksamkeit zu widmen. Dabei kommt der Schulung der Interviewer eine besondere Rolle zu. Zwar hat es sich als unerheblich erwiesen, ob mündlich oder schriftlich geschult wird. Um den Erfolg der Methode zu steigern, wird man allerdings zu einer differenzierteren Schulung übergehen müssen, bei der u.U. nicht mehr alle Interviewer ein und dasselbe Trainingsprogramm durchlaufen, und bei der bestimmte Interviewergruppen besonders sorgfältig auf die Aufgabe, die sie erwartet, vorbereitet werden. Darüber hinaus müssen die Interviewer mit den Schwierigkeiten, die sie bei bestimmten Befragtengruppen zu erwarten haben, vertraut gemacht werden. Ganz wesentlich dürfte hierbei auch sein, falsche Erwartungshaltungen bei den Interviewern abzubauen. Es hat sich ja z.B. gezeigt, daß es keineswegs so ist, daß ältere Befragte - wie vermutet wurde - mit der Methode die größten Schwierigkeiten haben. Bei Interviewern, die so denken, wird es bei dieser Befragtengruppe aber häufiger zu Ausfällen kommen.



Unabhängig von gezielten Schulungsverfahren bietet natürlich die wachsende Vertrautheit mit der neuen Methode die beste Gewähr dafür, daß sie auch angemessen durchgeführt wird. In dem Maße, in dem es zur gewohnten Aufgabe eines professionellen Interviewers wird, in einer Standardbefragung auch mit Magnitude-Skalierung zu arbeiten, dürften die Negativwirkungen bestimmter Kontexteffekte abnehmen - ganz abgesehen davon, daß dann auch der Schulungsaufwand geringer wird. Das heißt natürlich nicht, daß von irgendeinem Zeitpunkt an auf die Kontrolle kontextueller Effekte verzichtet werden sollte. Allerdings ist die Bestimmung der Qualität von Magnitude-Messungen ja ohnehin Bestandteil der Skalenkonstruktionen. Insofern ist jede Anwendung von Magnitude-Skalierung in der Forschung zugleich Skalierungsforschung.

## Anhang: Bibliographische Hinweise

Oberblicksreferate und die theoretische Rechtfertigung der Verwendung von Magnitude-Skalierung zur Messung von Einstellungen finden sich in dem Aufsatz A Metric for the Social Consensus von S.S. STEVENS (1966) sowie in Kapitel 8 seines posthum erschienenen Lehrbuches Psychophysics. Introduction to its Perceptual, Neural, and Social Prospects (1975, pp. 227-267). Nicht verwunderlich ist, daß beide Beiträge strikt von der Warte der "Stevens-Schule" der Psychophysik aus geschrieben sind, und alternativen Ansätzen dabei nur wenig Beachtung geschenkt wird. Als Gegengewicht und um den Stellenwert der STEVENSSchen Theorie innerhalb des Gesamtrahmens der Psychophysik abschätzen zu können, sind die Einführungslehrbücher z.B. von GESCHIEDER (1976) oder BAIRD & NOMA (1978) empfehlenswert. Eine Erörterung des Beziehungsverhältnisses von sensorischer Psychophysik und "Psychophysik der Einstellungen" unter strukturellen Gesichtspunkten ist bei WEGENER (1980a) zu finden.

In Hinsicht auf Untersuchungen, in denen Magnitude-Methoden zur Messung von Einstellungsvariablen eingesetzt werden, ist zwischen Laboratoriumsexperimenten und Feldstudien zu unterscheiden. Die Anzahl der Untersuchungen mit experimentellem Charakter ist unübersehbar groß, und entsprechend vielfältig sind auch die Einstellungsdimensionen, die man sich auf diese Weise zum Gegenstand macht (vgl. STEVENS, 1975). Ihre inhaltliche Relevanz für eine

sozialwissenschaftliche Theoriebildung liegt allerdings nicht immer auf der Hand. Für den vorliegenden Zusammenhang interessanter sind Studien, die Magnitude-Messungen in Felderhebungen vornehmen und deren Ziel die Verbesserung des Skalenniveaus von theoretisch wichtigen Einstellungsvariablen ist. Die Zahl solcher Untersuchungen ist jedoch nach wie vor nicht sehr groß. Darüber hinaus kommt die Magnitude-Skalierung häufig nur "monomodal" zum Einsatz - zumeist als numerische magnitude estimation -, wobei jedoch der besondere Vorteil, den die mehrmodale Anwendung durch die Überprüfbarkeit der Skalierungsgüte bietet, verschenkt wird.

Der erste Einsatz - monomodaler - Magnitude-Messung in einer Felderhebung wurde von HAMBLIN & SMITH (1966) vorgenommen. In dieser und in einer Reihe von Folgeuntersuchungen (HAMBLIN, 1971a, 1971b; SHINN, 1969a; RAINWATER, 1971; vgl. auch HAMBLIN, 1974) wurden Befragte aufgefordert, unterschiedliche Niveaus und Zusammensetzungen von Einkommen, Ausbildung und beruflicher Tätigkeiten mit Hilfe numerischer magnitude estimation einzuschätzen. Eine Zusammenfassung der inhaltlichen Ergebnisse und zusätzliches neues empirisches Material im Rahmen dieses Ansatzes finden sich in dem Buch Social Standing in America von COLEMAN & RAINWATER (1979). Weitere Untersuchungen mit Magnitude-Skalierungen, die nur eine Reaktionsmodalität verwenden, sind SHINNs Studien zur Wahrnehmung nationaler Macht (1969b) und zur Einschätzung der Wichtigkeit politischer Ämter (1969a; mit Linien-Modalität), SELLIN & WOLFGANGs umfangreiche Untersuchung über die Verwerflichkeit von Verbrechen und die Angemessenheit von Strafmaßen (1964) sowie CORSONs (1970) Arbeit zum nationalen Konflikt- und Kooperationsverhalten. WELCH (1971) konstruierte eine Magnitude-Skala, auf der das Ausmaß an Zufriedenheit mit dem amerikanischen politischen System gemessen wird. Eine Reihe beruflicher Einstellungen von Arbeitern einer holländischen Stahlfabrik wurden von SARIS et al. (1977) mit magnitude estimation erhoben.

Für die Anwendung mehrmodaler Magnitude-Messung in Befragungen hat sich als einfach handhabbare Reaktionsmodalität neben der Zahlenschätzung das Linienzeichnen bewährt (vgl. LODGE et al., 1976a; WEGENER, 1978, 1979). Es gibt allerdings auch Versuche, die Skalierungen mit (transportablen) Apparaturen für die Manipulation von Lautstärken und Handdruck durchzuführen. Auf diese Weise haben etwa LODGE et al. (1975, 1976b; vgl. auch LODGE & TURSKY, 1980) die politische Unterstützung gemessen, die verschiedenen politischen Maßnah-

men und Persönlichkeiten entgegengebracht wird. Allerdings sind die Einsatzmöglichkeiten eines solchen Vorgehens aus offensichtlichen Gründen begrenzt. In Normalbefragungen kommt daher nur das Papier-und-Bleistift-Verfahren mit Zahlen und Linien in Frage. Bis jetzt liegen die folgenden Untersuchungen vor, in denen diese Art der doppelten Magnitude-Skalierung eingesetzt wurde: In einer Repräsentativerhebung haben MURPHY & TANENHAUS (1974) die politische Unterstützung für den U.S. Supreme Court gemessen. LODGE & TURSKY (1979) benutzten das mehrmodale Magnitude-Verfahren zur Einstellungsmessung in einer regionalen Wahluntersuchung; dabei wurden die standardisierten Frageformulierungen des Survey Research Center, Michigan, zugrundegelegt, um Vergleiche mit konventionellen Abfragetechniken zu ermöglichen. Die erste deutsche Untersuchung wurde an einer Lehrlingspopulation vorgenommen (BECK et al., 1979) und setzte die Magnitude-Methode bei der Erhebung beruflicher Einstellungen und von Einstellungen zur Berufswahl ein. Außerdem liegt der Bericht über eine Pilotstudie vor (GRUNERT, 1980), in der subjektive Kaufrisiken magnitude-skaliert wurden.

Ein für die Benutzung von Magnitude-Messungen wichtiges Thema ist die Frage nach der Vergleichbarkeit von Magnitude- mit kategorischen Skalen. Im Rahmen der reinen Psychophysik hat sich die Untersuchung von Interskalenrelationen als ein Angelpunkt für die Bestimmung von Skaleneigenschaften erwiesen (KRANTZ, 1972; EISLER et al., 1979; MONTGOMERY, 1980; ORTH, 1980). Darüber hinaus basieren neuere Ansätze zur Validierung von kognitiven Integrationsmodellen auf einer Gegenüberstellung von Empfindungsskalen, die mit unterschiedlichen Methoden konstruiert wurden (vgl. ANDERSON, 1980; BIRNBAUM, 1980). Für die Einstellungsmessung in Umfragen steht allerdings zunächst ein bescheideneres Ziel im Vordergrund: Die Bestimmung der funktionalen Beziehung zwischen den unterschiedlichen Skalentypen ist für sich genommen erstrebenswert, weil nur dann, wenn eine solche, generalisierbare Übersetzungsfunktion bekannt ist, Ergebnisse, die mit unterschiedlichen Meßverfahren gewonnen werden, miteinander vergleichbar sind. Der erste Versuch, Kommensurabilitätsbedingungen für Einstellungsskalen unterschiedlichen Typs zu formulieren, wurde von SHINN (1974) unternommen. Die bereits erwähnte Arbeit von LODGE & TURSKY (1979) stellt kategorische und Magnitude-Skalen in Bezug auf typische politikwissenschaftliche Einstellungsvariablen gegenüber. Auf der Basis schließlich von 12 kategorisch und mit Magnitude-Technik gemessenen Einstellungsskalen aus Umfragen schlägt WEGENER (1980b)

ein Interskalenmodell vor, in das als Parameter auch individuelle Kontextbedingungen der beiden beteiligten Messungen eingehen. WEGENER & KIRSCHNER (1980) erläutern das mathematische Schätzverfahren, das zur Bestimmung entsprechend komplexer Interskalenfunktionen anzuwenden ist.

Abschließend ist noch auf die Auswertung von Magnitude-Daten einzugehen, d.h. auf das technische Verfahren zur Konstruktion von Skalenwerten aus zwei oder mehr Meßreihen unterschiedlicher Reaktionsmodalitäten. Bei ZUMA liegt hierfür - als Bestandteil des ZUMAPACKs - das Programm ZUMAPSYCH vor, das im dazugehörigen Manual erläutert wird (vgl. auch WEGENER, 1978). Die Arbeit von CROSS (1974) behandelt wesentliche Aspekte des mathematischen Hintergrunds des Rechenprogramms.

- ANDERSON, N.H. Cognitive algebra and social psychophysics. In: WEGENER, B. (1980), 1980.
- BAIRD, J.C. & NOMA, E. Fundamentals of scaling and psychophysics. New York: Wiley, 1978.
- BECK, U., BRATER, M. & WEGENER, B. Berufswahl und Berufszuweisung. Zur sozialen Verwandtschaft von Ausbildungsberufen. Frankfurt: Campus, 1979.
- BIRNBAUM, M.H. Controversies in psychological measurement. In: WEGENER, B. (1980), 1980.
- COLEMAN, R.P. & RAINWATER, L. Social standing in America. London: Routledge & Kegan Paul, 1979.
- CORSON, W.H. Conflict and cooperation in East-West crises: Dynamics of crisis interaction. Ph.D. dissertation, Havard University, Dept. of Sociology, 1970.
- CROSS, D.V. Some technical notes on psychophysical scaling. In: MOSKOWITZ, H., SCHARF, B. & STEVENS, J.C. (Eds.). Sensation and measurement: Papers in honor of S.S. Stevens. Dordrecht: Reidel, 1974, 23-36.
- EISLER, H., HOLM, S. & MONTGOMERY, H. The general psychophysical differential equation: A comparison of three specifications. Journ. Math. Psychol., 20, 1979, 16-34.
- GESCHEIDER, G.A. Psychophysics: Method and theory. Hillsdale, N.J.: Erlbaum, 1976.
- GRUNERT, K.G. Die Eignung des Magnitude-Verfahrens zur Skalierung der Intensität des Empfindens von Kaufrisiken. Arbeitspapier 9. Strategien der

- Verbraucherinformation. Vergleichende Analyse der Effizienz informations-politischer Maßnahmen, Hohenheim, 1980.
- HAMBLIN, R.L. Mathematical experimentation and sociological theory: A critical analysis. Sociometry, 34, 1971a, 423-452.
- HAMBLIN, R.L. Ratio measurement for the social sciences. Social Forces, 50, 1971b, 191-206.
- HAMBLIN, R.L. Social attitudes: Magnitude measurement and theory. In: BLA-LOCK, H.M. (Ed.). Measurement in the social sciences: Theories and strategies. Chicago: Aldine, 1974, 61-120.
- HAMBLIN, R.L. & SMITH, C.R. Values, status, and professors. Sociometry, 29, 1966, 183-196.
- KRANTZ, D.H. Magnitude estimation and cross-modality matching. Journ. Math. Psychol., 9, 1972, 168-199.
- LODGE, M. & TURSKY, B. Comparison between category and magnitude scaling of political opinion employing SRC/CPS items. American Political Review, 73, 1979, 50-66.
- LODGE, M. & TURSKY, B. The social-psychological scaling of political opinion. In: WEGENER, B. (1980), 1980.
- LODGE, M., CROSS, D.V., TURSKY, B. & TANENHAUS, J. The psychophysical scaling and validation of a political support scale. Am. Journ. Pol. Sci., 1975, 19, 611-649.
- LODGE, M., CROSS, D.V., TURSKY, B., TANENHAUS, J. & REEDER, R. The psychophysical scaling of political support in the "real world". Pol. Methodology, 1976a, 2, 159-182.
- LODGE, M., CROSS, D.V., TURSKY, B., FOLEY, M.A. & FOLEY, H. The calibration and cross-modal validation of ratio scales of political opinion in survey research. Soc.Sci.Research, 1976b, 5, 352-347.
- MONTGOMERY, H. Intra- and interindividual variations in the form of psychophysical scales. In: WEGENER, B. (1980), 1980.
- MURPHY, W. & TANENHAUS, J. Explaining diffuse support for the United States Supreme Court: An assessment of four models. Notre Dame Lawyer, 1974, 49, 1037-1044.
- ORTH, B. A theoretical and empirical study of scale properties of magnitude estimation and category rating scales. In: WEGENER, B. (1980), 1980.
- RAINWATER, L. The measurement of social status. Ph.D. dissertation, Harvard University, Dept. of Sociology, 1971.
- SARIS, W.E., BRUINSMA, C., SHOOTS, W. & VERMEULEN, C. The use of magnitude

- estimation in large scale survey research. Mens en Maatschappij, 52, 1977, 369-395.
- SELLIN, J.T. & WOLFGANG, M.E. The measurement of delinquency. New York: Wiley, 1964.
- SHINN, A.M. The application of psychophysical scaling techniques to measurement of political variables. Institute for Research in Social Science, University of North Carolina: Chapel-Hill Press, 1969a.
- SHINN, A.M. An application of psychophysical scaling techniques to the measurement of national power. Journal of Politics, 31, 1969b, 932-951.
- SHINN, A.M. Relations between scales. In: BLALOCK, H.M. (Ed.). Measurement in the social sciences. Theories and strategies, Chicago: Aldine, 1974, 121-158.
- STEVENS, S.S. On the brightness of lights and the loudness of sounds. Science, 118, 1953, 576 (Abstract).
- STEVENS, S.S. A metric for the social consensus. Science, 151, 1966, 530-541.
- STEVENS, S.S. Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: Wiley, 1975.
- WEGENER, B. Einstellungsmessung in Umfragen: Kategorische vs. Magnitude-Skalen. Zumanachrichten, 3, 1978, 3-27.
- WEGENER, B. Magnitude-Messung beruflicher Einstellungen. In: BECK, U., BRATER, M. & WEGENER, B. Berufswahl und Berufszuweisung. Zur sozialen Verwandtschaft von Ausbildungsberufen. Frankfurt: Campus, 1979, 125-162.
- WEGENER, B. (Ed.) Social attitudes and psychophysical measurement. Hillsdale, N.J.: Erlbaum, 1980 (im Druck).
- WEGENER, B. Outline of a structural taxonomy of sensory and social psychophysics. In: WEGENER, B. (1980), 1980a.
- WEGENER, B. Fitting category to magnitude scales for a dozen survey-assessed attitudes. In: WEGENER, B. (1980), 1980b.
- WEGENER, B. & KIRSCHNER, H.-P. A note on estimating interscale relationships in "direct" psychophysical scaling, 1980 (zur Veröffentlichung eingereicht).
- WELCH, R.E. The use of magnitude estimation in attitude scaling: Constructing a measure of political dissatisfaction. Social Science Quarterly, 1971, 76-87.
- ZUMAPACK Manual. Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, 1980.